THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 263, No. 29, Issue of October 15, pp. 14739-14747, 1988
Printed in U.S.A.

Structure of the Human Neutrophil Elastase Gene*

(Received for publication, March 24, 1988)

Hideki Takahashi, Toshihiro Nukiwa, Kunihiko Yoshimura, Caryn D. Quick, David J. States, Mark D. Holmes, Jacqueline Whang-Peng‡, Turid Knutsen‡, and Ronald G. Crystal§

From the Pulmonary Branch, National Heart, Lung, and Blood Institute, and the ‡Medicine Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892

The gene for human neutrophil elastase (NE), a powerful serine protease carried by blood neutrophils and capable of destroying most connective tissue proteins, was cloned from a genomic DNA library of a normal individual. The NE gene consists of 5 exons and 4 introns included in a single copy 4-kilobase segment of chromosome 11 at q14. The coding exons of the NE gene predict a primary translation product of 267 residues including a 29-residue N-terminal precursor peptide and a 20-residue C-terminal precursor peptide. Analysis of the N-terminal peptide sequence suggests it contains a 27-residue "pre" signal peptide followed by a "pro$_N$" dipeptide, similar to that of other blood cell lysosomal proteases. The sequences for the mature 218-residue NE protein are included in exons II–V. The 5′-flanking region of the gene includes typical TATA, CAAT, and GC sequences within 61 base pairs (bp) of the cap site. The sequence 1.5 kilobases 5′ to exon I contains several interesting repetitive sequences including six tandem repeats of unique 52- or 53-bp sequences. The 5′-flanking region also contains a 19-bp segment with 90% homology to a segment of the 5′-flanking region of the human myeloperoxidase (MPO) gene, a gene also expressed in bone marrow precursor cells and a protein stored in the same neutrophil granules as NE. In addition, like the MPO gene, the NE 5′-flanking region has several regions with ≥75% homology to sequences 5′ to c-*myc*, but there is no overlap between the NE-c-*myc* and MPO-c-*myc* homologous sequences.

Human neutrophil elastase (NE[1]; EC 3.4.21.37), a 218-amino acid glycoprotein with two asparaginyl *N*-linked carbohydrate side chains and four intramolecular disulfide bridges, functions as a powerful serine protease capable of cleaving most protein components of the extracellular matrix, a variety of proteins of the coagulation and complement cascades, and *Escherichia coli* cell wall components (1–5). NE is classed as an elastase because it is one of a small group of mammalian proteases that function at neutral pH to cleave elastin, a highly cross-linked rubber-like macromolecule present in many extracellular matrices (3, 5). The sequence of the mature NE protein has been determined (2) as has the three-dimensional structure using x-ray crystallography (4). Like other proteases in its class, NE function is dependent on the catalytic triad His^{41}-Asp^{88}-Ser^{173} centered at the NE reactive site (3, 4, 6). When a substrate is presented within the NE active site pocket, the transfer of a proton within the triad allows the Ser^{173} to become a highly reactive nucleophile capable of attacking the peptide bond within the target substrate (4, 6).

The bulk of NE in the human body is found in the mature neutrophil where it is stored in azurophilic granules, lysosome-like structures distributed throughout the cytoplasm (3). Despite its name, NE is not produced by neutrophils, only carried by them (7). The available evidence suggests that the human NE gene is only expressed in bone marrow myelocytic progenitor cells, and the gene is turned off prior to the time neutrophils leave the marrow (7). The mature neutrophils utilize the stored enzyme in phagolysosomes or export NE when the neutrophil is stimulated or has been lysed (3, 5).

In addition to its likely role in normal tissue turnover and host defense, NE is thought to play a major role in tissue damage in acute and chronic inflammatory disorders (3, 5). Abnormalities in NE gene expression have been implicated in the pathogenesis of the susceptibility to infections in the hereditary Chediak-Higashi syndrome (8). Furthermore, α_1-antitrypsin deficiency, a hereditary disorder associated with inadequate defenses against NE, results in chronic destruction of the alveolar wall and the development of emphysema (9).

In the context of these observations, it is apparent that knowledge of the structure and function of the NE gene has major implications for health and disease. As an initial step toward understanding NE gene expression, the purpose of the present study is to characterize the structure of the gene for this powerful serine protease.

MATERIALS AND METHODS

Cloning and Sequencing of the Neutrophil Elastase Gene—A DNA library from an incomplete *Sau*3A digest of genomic DNA of a normal individual was constructed in λ phage EMBL3 using standard methods (10). Starting with 5 × 10^5 λ phage plaque-forming units, the library was screened using a ^{32}P-labeled 0.65-kb NE cDNA clone (pPB15) insert as a probe (7). Three positive plaques were identified. One 18-kb clone, designated λNE18.0, was chosen for study.

The DNA sequencing of the exons, exon-intron junctions, and the 5′- and 3′-flanking regions of the NE gene was carried out by the dideoxynucleotide chain termination procedure (11) using modified T7 DNA polymerase (United States Biochemical Corp.) (12) and synthetic bidirectional oligonucleotide primers (10). To accomplish this, the phage clone λNE18.0 was subcloned using M13 phage vectors mp18 or mp19 or the plasmid vectors pUC13 or pUC19.

Identification and sequencing of the coding exons of the NE were aided by knowledge of the protein sequence of the 218 residues of the mature NE protein (2) together with the DNA sequence of the partial NE cDNA clone (pPB15) that codes for residues 46–218 of the mature protein plus a putative 20-amino acid C-terminal "precursor" peptide

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J04032.

§ To whom reprint requests should be addressed: Pulmonary Branch, Bldg. 10, Rm. 6D03, NIH, Bethesda, MD 20892.

[1] NE, neutrophil elastase; kb, kilobase(s); bp, base pair(s); MPO gene, myeloperoxidase gene.

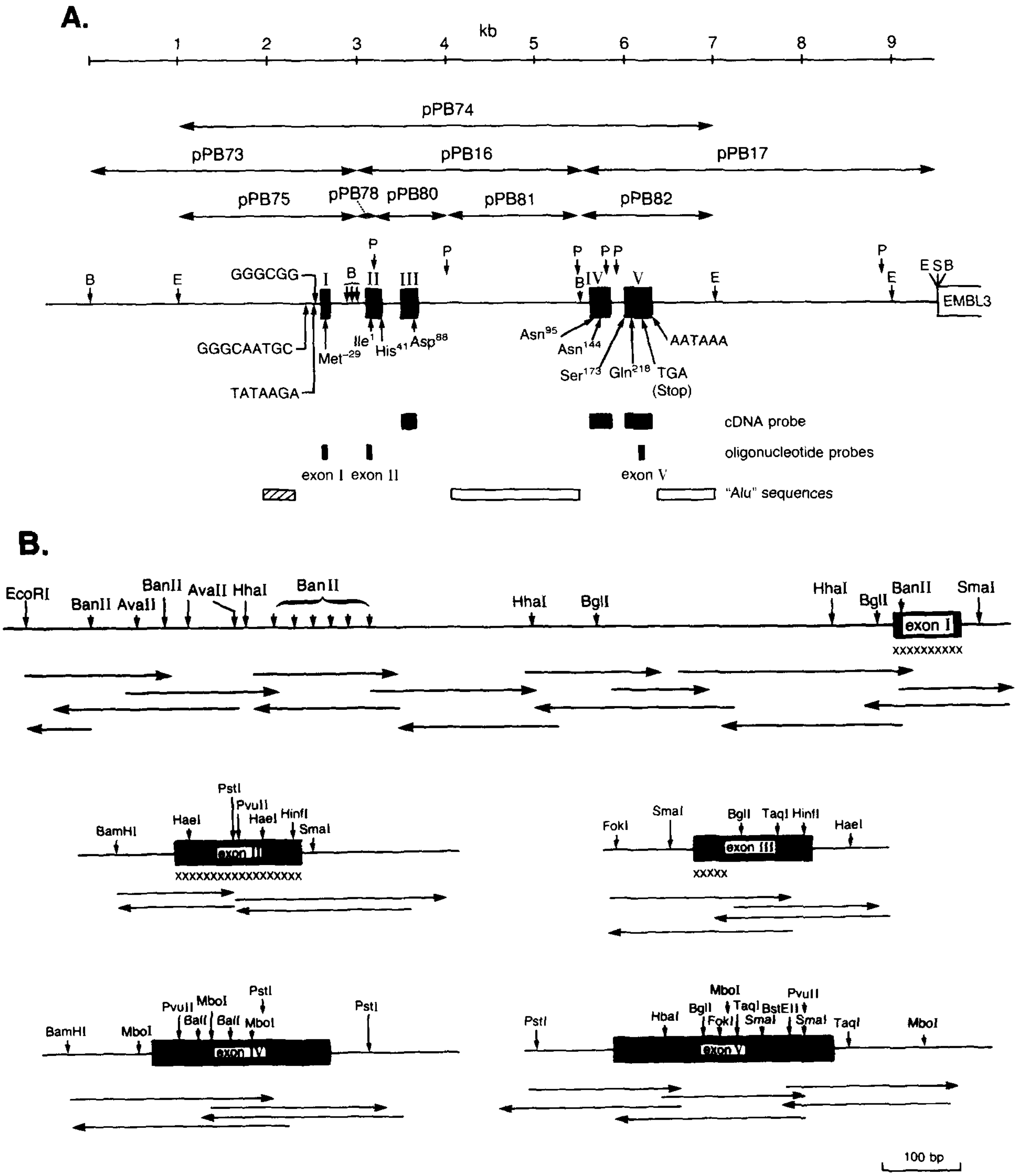

FIG. 1. **Structure of the human neutrophil elastase (NE) gene.** Exons are depicted by *solid boxes* (labeled *I–V*), introns and flanking regions by a *thin line*. *A*, overview of the NE gene including five exons, promoter consensus sequences (GGGCAATGC, TATAAGA, GGGCGG) 5′ to exon I, the start codon (ATG, Met^{-29}) in exon I, and the stop codon (TGA) and putative polyadenylation signal (AATAAA) in exon V. The codons for N-terminal (Ile^{1}) and C-terminal (Gln^{218}) amino acid residues of mature NE protein reside in exon II and exon V, respectively. The codons for the "catalytic triad" His^{41}-Asp^{88}-Ser^{173}, which forms the reactive site of NE, reside in exons II, III, and V, respectively. The codons for putative *N*-linked glycosylation sites Asn^{95} and Asn^{144} are in exon IV. Restriction enzyme sites indicated include *Bam*HI (*B*), *Eco*RI (*E*), *Pst*I (*P*), and *Sal*I (*S*). The *box* at the 3′ end represents the EMBL3 λ phage arm. *Horizontal arrows above* the restriction map represent genomic subclones in pUC13 (pPB16, pPB74), pUC19 (pPB17, pPB 73), or M13 (pPB75, pPB78, pPB80-pPB82). The regions marked *cDNA probe* below the restriction map are the portions of the exon encompassed by the cDNA clone (pPB15) used for the cloning of NE gene. The oligonucleotide gene probes used to detect exons I, II, and V are also indicated. Three regions of Alu sequences are present; those indicated by *open boxes* define regions 3′ to exon I detected by hybridization to DNA sequences of the Alu type (BLUR 8), and that indicated by a *hatched box* defines an Alu sequence confirmed by sequencing. *B*, detailed restriction map and sequencing strategy for 1.5 kb 5′ to exon I, the five exons with portions of the adjacent introns, and the 3′-flanking regions of the NE gene. *Horizontal arrows* indicate the direction and length of each fragment of DNA sequenced; X-marked regions also are sequenced at the RNA level (see text).

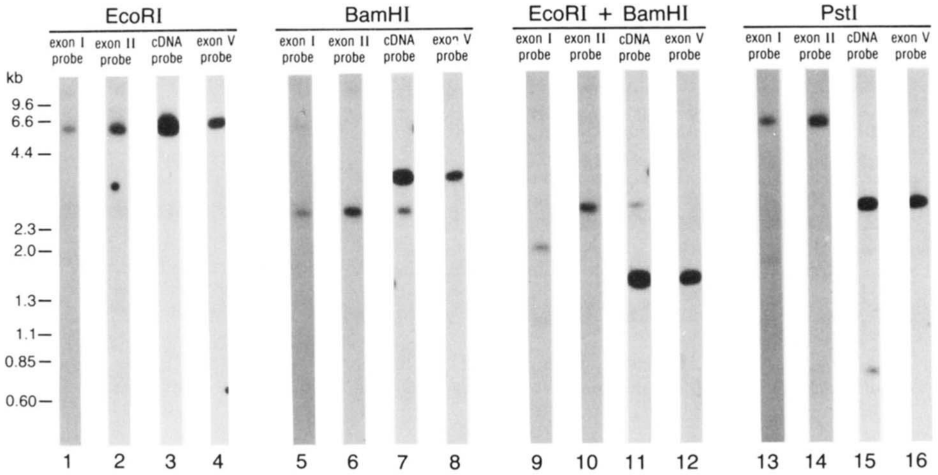

FIG. 2. **Southern analysis of neutrophil elastase genomic clone λNE18.0, representing 18 kb of genomic DNA encompassing the entire NE gene along with 12-kb 5′-flanking and 2-kb 3′-flanking regions.** The clone was digested with restriction endonuclease *Eco*RI (*lanes 1–4*), *Bam*HI (*lanes 5–8*), *Eco*RI + *Bam*HI (*lanes 9–12*), and *Pst*I (*lanes 13–16*) and evaluated by Southern analysis followed by hybridization with ^{32}P-labeled cDNA probe (*lanes 3, 7, 11, 15*), exon I oligonucleotide probe (*lanes 1, 5, 9, 13*), exon II probe (*lanes 2, 6, 10, 14*), and exon V probe (*lanes 4, 8, 12, 16*) as indicated. See Fig. 1 for details. All exons are included in the 6.0-kb *Eco*RI fragment.

(7). To establish the exon sequences for those exon regions 5′ to residue 46 of the mature protein, RNA sequencing was carried out by the method of Geliebter *et al.* (13) using poly(A)$^+$ NE mRNA isolated from the human promyelocytic cell line HL60, known to express NE mRNA (7).

Overall Structure of the Neutrophil Elastase Gene—The exon-intron structure of the genomic clone λNE18.0 was determined from the genomic DNA sequence, partial cDNA sequence (7), RNA sequence, restriction endonuclease mapping, and primer extension analysis (14). The probes used to map the gene included the cDNA clone (pPB15) together with 17-mer oligonucleotide probes (see Fig. 1*A* for identification of the regions covered by the individual probes). The cDNA was labeled with ^{32}P by nick translation (15) and the oligonucleotide probes by 5′ end labeling (16). The location of the 5′ end of the neutrophil elastase mRNA was identified by primer extension analysis (14) using a ^{32}P 5′ end-labeled oligonucleotide (corresponding to residues 13–19 of the mature protein) as a primer.

During the evaluation of the overall structure of the NE gene, Southern analysis of human total genomic DNA with the NE gene subclones pPB16 and pPB17 as probes revealed a pattern of hybridization suggesting that the subclones contained sequences which were present in high copy number in genomic DNA. To evaluate the possibility that some of these sequences represented the common "Alu" family of repetitive gene sequences (17), the NE genomic subclone pPB74 was digested with restriction endonucleases and evaluated by Southern analysis using a ^{32}P-labeled DNA sequence of the Alu type (BLUR 8) as a probe (17).

Homologies with Other Genes—Homology search of the GenBank™ and National Biomedical Research Foundation data bases were performed using the DNASIS and PROSIS software from Hitachi America Ltd.

Copy Number and Chromosomal Localization—The number of NE gene copies in normal human genomic DNA was determined using ^{32}P-labeled NE cDNA pPB15 as a probe and DNA (10 μg) from the NE genomic clone λNE18.0 as a standard. To accomplish this, a standard curve was constructed using known amounts of *Eco*RI-digested λNE18.0 DNA equivalent to 0, 0.5, 1.0, 1.5, or 2.0 gene copies with *Micrococcus lysodeikticus* DNA (10 μg, *Eco*RI-digested) as a carrier (18). The samples were applied in individual dots on nitrocellulose in parallel with 10-μg aliquots of *Eco*RI-digested human genomic DNA, and hybridization was carried out with the ^{32}P-labeled cDNA probe. Quantitation was accomplished by scanning the autoradiograms.

The chromosomal localization of the NE gene was determined by *in situ* hybridization using the NE cDNA (pPB15) as a probe (7). Human metaphase and prometaphase chromosome spreads were performed from fluorodeoxyuridine-synchronized cultures of normal human blood lymphocytes. After denaturing the chromosomal DNA, *in situ* hybridization was carried out at 37 °C for 16 h using a ^{3}H-labeled cDNA probe (specific activity, 10^7 dpm/μg). Autoradiograms of the hybridized slides were prepared and stored at 4 °C for 22 days, and G-banding of the chromosomes was accomplished using 0.25% Wright stain for 5 min. Grain locations on the autoradiographs were determined (19).

RESULTS

Evaluation of the human NE genomic clone λNE18.0 by DNA sequencing, partial RNA sequencing, Southern analysis and primer extension analysis revealed 5 exons and 4 introns dispersed over a 4-kb region (Figs. 1–4). The 5 exons range in length from 93 bp (exon I) to more than 270 bp (exon V). The smallest intervening sequence is between exons IV and V (approximately 170 bp), while the largest is between exons III and IV (approximately 1.8 kb). The major initiation point of transcription is located 26 bases 5′ to the ATG start codon (Met^{-29}, Fig. 4). The primer extension analysis also demonstrated a faint band approximately 90 bp 5′ from this ATG; it is possible this presents a minor site of transcription initiation. The single, in-frame start codon for translation (Met^{-29}) is located in exon I, and the single, in-frame stop codon (TGA) is located in exon V. Within 61 bp of exon I, the 5′ portion of the gene is flanked by consensus promoter elements, including a CAAT box (GGGCAATGC, −61 to −53), a TATA box flanked with G+C-rich segments (TATAAGA, −31 to −25), and a GC box (GGGCGG, −18 to −13). No "classic" promoter elements were identified 5′ to the putative minor cap site approximately 90 bp from the start ATG, but there is a CAAT-like box in the region starting at −150 and a TATA-like box in the region starting at −624. In exon V, 58 bp 3′ to the stop codon there is a typical polyadenylation signal sequence (AATAAA).

The protein sequence derived from the genomic DNA sequence predicts that NE is synthesized as a precursor protein of 267 amino acids, including a 29-residue N-terminal extension and a 20-residue C-terminal extension flanking the 218-residue mature protein (Fig. 3). The 29-residue peptide N-terminal to the Ile^1 of the mature protein contains a high proportion of hydrophobic amino acids, typical of a signal sequence (20, 21). A typical consensus signal peptidase cleavage signal (Ala-Leu-Ala) (22) ends at Ala^{-3}, suggesting that the N-terminal extension actually represents a 27-residue "pre" signal peptide (Met^{-29} → Ala^{-3}) followed by a 2-residue "pro$_N$" peptide (Ser^{-2}-Glu^{-1}).

Following the "pre-pro$_N$" extension, the next 218 residues predicted by the genomic sequence are identical to the sequence of the mature neutrophil elastase protein as determined by Sinha *et al.* (2). Comparison of the genomic DNA sequence to the partial cDNA sequence of clone pPB15 derived from the tumor cell line U937 (7) demonstrated that they were identical except for a single, synonymous base difference in the codon for Ser^{173} (Fig. 3). Interestingly, comparison of the sequence of the human neutrophil elastase protein, partial cDNA sequence, and genomic DNA sequence

```
                                                                                 -1418 ···GAATTCTCTCTCCAGCAG
                                             +
-1400   CCCTGCCAGATGCCCGCCCAGCCCCTGCCTCAGGCGGGGAGGGCTTCAGGGAAGCTCACCAAGGCAGAAGGGCGGGAGAGATTGTCAGAGCCCCAGCTGG
                                                      +
-1300   TGTCCAGGGACTGACCGTGAGCCTGGGTGAAAGTGAGTTCCCCGTTGGAGGCAACAGACGAGGAGAGGATGGAAGGCCTGGCCCCCAAGAATGAGCCCTG
                                                                                                 +
-1200   AGGTTCAGGGAGCGGCTGGAGTGAGCCGGCCCCAGATCTCCGTCCAGCTGCGGGTCCCAGAGGCCTGGGTTACACTCGCAGCTCCTGGGGGAGGCCCTTG
                                                                            ♦
-1100   ACGTGCCTCAGTTCCCAAACAGGAACCCTGGGAAGGACCAGAGAAGTGCCTATTGCGCAGTGAGTGCCCGACACAGCTGCATGTGGCCGGTATCACAGGG
                   ●         ♦                                          ●         ♦
-1000   CCCTGGGTAAACTGAGGCAGGCGACACAGCTGCATGTGGCCGGTATCACAGGGCCCTGGGTAAACTGAGGCAGGCGACACAGCTGCATGTGGCCGGTATC
                         ●         ♦                                         ●         ♦
 -900   ACAGGGCCCTGGGTAAACTGAGGCAGGCGACACAGCTGCATGTGGCCGTATCACAGGGCCCTGGGTAAACTGAGGCAGGTGACACAGCTGCATGTGGCCG
                              ●         ♦                                          ●
 -800   GTATCACGGGGCCCTGGATAAACAGAGGCAGGCGACACAGCTGCATGTGGCCGGTATCACGGGGCCCTGGGTAAACTGAGGCAGGCGAGGCCACCCCCAT

 -700   CAAGTCCCTCAGGTCTAGGTTTGGCAGGTTTGGCAAAAACACAGCAACGCTCGGTTAAATCTGAATTTCGGGTAAGTATATCCTGGCCTCATTTGGAAG

 -600   AGACTTAGATTAAAAAAAAAACGTCGAGACCAGCCCGGCCAACACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCAGGCGCAGTGGCTCA
                              +    ●
 -500   CGCCTGTGATCCCAGCACTCTGGGAGGCTGAGGCAGGCGGATCACCCGAGGTCAGATGTTCAAGACCAGCCTGGCCGACAGGGCGAAACACTGTCTCTAC
                                                                 +    ●                          +     +
 -400   TACAAATACAAAAATTAGCCGGGAGTGGTGGCAGGTGCCTGTAATCTCAGCTATTCAGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCGGAGG

 -300   TTGCCGTGAGCCGGGATCACGCCACCGCACTCCAGCCTGGGCGATAGAGCAAGACTCTGTCTCCAAAAAAATAAATTAAAAAACCCACATTGATTATCTG
                                            +
 -200   ACATTTGAATGCGATTGTGCATCCTGAATTTTGTCTGGAGGCCCCACCCGAGCCAATCCAGCGTCTTGTCCCCCTTCTCCCCCTTTTCATCAACGCCCTG
                          +                                                                         +
 -100   TGCCAGGGGAGAGGAAGTGGAGGGCGCTGGCCGGCCGTGGGGCAATGCAACGGCCTCCCAGCACAGGGCTATAAGAGGAGCCGGGCGGGCACGGAGGGGC

   ←exon I                          start
                                    ▼
+1 AGAGACCCCGGAGCCCCAGCCCCACC       ATG   ACC   CTC   GGC   CGC   CGA   CTC   GCG   TGT   CTT   TTC   CTC   GCC   TGT   GTC   CTG   CCG   GCC
                                    Met   Thr   Leu   Gly   Arg   Arg   Leu   Ala   Cys   Leu   Phe   Leu   Ala   Cys   Val   Leu   Pro   Ala
                                    -29                     -25                           -20                           -15               -12

                    exon I→                                                                                                                  ←exon II
TTG   CTG   CTG   GGG   Ggt   gagttttgagtccaacctcccgctgctccctctgtcccgggttctgt···   ···ggacaggctccttggcaggcactcagcacccgcacccggtgtgtcccca   gGC   ACC
Leu   Leu   Leu   Gly   (Gly)                                                                                                             (Gly) Thr
-11   -10               -7                                                                                                                -7    -6

GCG   CTG   GCC   TCG   GAG   ATT   GTG   GGG   GGC   CGG   CGA   GCG   CGG   CCC   CAC   GCG   TGG   CCC   TTC   ATG   GTG   TCC   CTG   CAG   CTG
Ala   Leu   Ala   Ser   Glu   Ile   Val   Gly   Gly   Arg   Arg   Ala   Arg   Pro   His   Ala   Trp   Pro   Phe   Met   Val   Ser   Leu   Gln   Leu
-5                      -1    1                       5                             10                            15                            20

CGC   GGA   GGC   CAC   TTC   TGC   GGC   GCC   ACC   CTG   ATT   GCG   CCC   AAC   TTC   GTC   ATG   TCG   GCC   GCG   CAC   TGC   GTG   GCG   AAT
Arg   Gly   Gly   His   Phe   Cys   Gly   Ala   Thr   Leu   Ile   Ala   Pro   Asn   Phe   Val   Met   Ser   Ala   Ala   His   Cys   Val   Ala   Asn
21                      25                            30                            35                            40                            45

exon II→                                                                                                         ←exon III
GTg   taagtagccgggagtgtgcgcgcccggctcggcatctcgcacgccgcgt···   ···gccccgaccccggggccgcccctgagccccgcctctccctccccggc   agA   AAC   GTC   CGC   GCG
(Val)                                                                                                             (Val) Asn   Val   Arg   Ala
46                                                                                                                46                      50

GTG   CGG   GTG   GTC   CTG   GGA   GCC   CAT   AAC   CTC   TCG   CGG   CGG   GAG   CCC   ACC   CGG   CAG   GTG   TTC   GCC   GTG   CAG   CGC   ATC
Val   Arg   Val   Val   Leu   Gly   Ala   His   Asn   Leu   Ser   Arg   Arg   Glu   Pro   Thr   Arg   Gln   Val   Phe   Ala   Val   Gln   Arg   Ile
51                      55                            60                            65                            70                            75

                                                                                                exon III→
TTC   GAA   AAC   GGC   TAC   GAC   CCC   GTA   AAC   TTG   CTC   AAC   GAC   ATC   GTG   ATT   CTC   CAG gtgccgccgggcgggcgggggcgcaggggcggaggcaga
Phe   Glu   Asn   Gly   Tyr   Asp   Pro   Val   Asn   Leu   Leu   Asn   Asp   Ile   Val   Ile   Leu   Gln
76                      80                            85                            90                93

                                                                       ←exon IV
gcgtggggagg···   ···atcatcactgccccgtgtgacgcgctgacgatctgtccccaccgccacag  CTC   AAC   GGG   TCG   GCC   ACC   ATC   AAC   GCC   AAC   GTG   CAG   GTG
                                                                        Leu   Asn   Gly   Ser   Ala   Thr   Ile   Asn   Ala   Asn   Val   Gln   Val
                                                                        94    95                            100                           105   106

GCC   CAG   CTG   CCG   GCT   CAG   GGA   CGC   CGC   CTG   GGC   AAC   GGG   GTG   CAG   TGC   CTG   GCC   ATG   GGC   TGG   GGC   CTT   CTG   GGC
Ala   Gln   Leu   Pro   Ala   Gln   Gly   Arg   Arg   Leu   Gly   Asn   Gly   Val   Gln   Cys   Leu   Ala   Met   Gly   Trp   Gly   Leu   Leu   Gly
107               110                           115                           120                           125                           130   131

AGG   AAC   CGT   GGG   ATC   GCC   AGC   GTC   CTG   CAG   GAG   CTC   AAC   GTG   ACG   GTG   GTG   ACG   TCC   CTC   TGC   CGT   CGC   AGC   AAC
Arg   Asn   Arg   Gly   Ile   Ala   Ser   Val   Leu   Gln   Glu   Leu   Asn   Val   Thr   Val   Val   Thr   Ser   Leu   Cys   Arg   Arg   Ser   Asn
132               135                           140                           145                           150                           155   156

                                                                        exon IV→
GTC   TGC   ACT   CTC   GTG   AGG   GGC   CGG   CAG   GCC   GGC   GTC   TGT   TTC gtacgtgccctgggtgtccctctgctccccacccgctcccagcccggact···   ···ggtggg
Val   Cys   Thr   Leu   Val   Arg   Gly   Arg   Gln   Ala   Gly   Val   Cys   Phe
157               160                           165                           170

                                                ←exon V      ▼
cagggcctcgcagtccagcttccccaccttgtctgcctccacag    GGG   GAC   TCC   GGC   AGC   CCC   TTG   GTC   TGC   AAC   GGG   CTA   ATC   CAC   GGA   ATT   GCC   TCC   TTC
                                                Gly   Asp   Ser   Gly   Ser   Pro   Leu   Val   Cys   Asn   Gly   Leu   Ile   His   Gly   Ile   Ala   Ser   Phe
                                                171                     175                           180                           185                     189

GTC   CGG   GGA   GGC   TGC   GCC   TCA   GGG   CTC   TAC   CCC   GAT   GCC   TTT   GCC   CCG   GTG   GCA   CAG   TTT   GTA   AAC   TGG   ATC   GAC
Val   Arg   Gly   Gly   Cys   Ala   Ser   Gly   Leu   Tyr   Pro   Asp   Ala   Phe   Ala   Pro   Val   Ala   Gln   Phe   Val   Asn   Trp   Ile   Asp
190                           195                           200                           205                           210                     214

TCT   ATC   ATC   CAA   CGC   TCC   GAG   GAC   AAC   CCC   TGT   CCC   CAC   CCC   CGG   GAC   CCG   GAC   CCG   GCC   AGC   AGG   ACC   CAC   TGA
Ser   Ile   Ile   Gln   Arg   Ser   Glu   Asp   Asn   Pro   Cys   Pro   His   Pro   Arg   Asp   Pro   Asp   Pro   Ala   Ser   Arg   Thr   His   ▼
215               218         220                           225                           230                           235               238   stop

GAAGGGCTGCCCGGGTCACCTCAGCTGCCCACACCCACACTCTCCAGCATCTGGCACAATAAACATTCTCTGTTTTGT···
```

FIG. 3. **Sequence of the 5′-flanking regions, all 5 exons and exon-intron junctions, and the 3′-flanking region of the neutrophil elastase gene.** The nucleotide sequences (*first line*) are shown with the deduced amino acid sequence (*second line*). The region 5′ to exon I, all exons, and the region 3′ to exon V are in *upper case letters* while the intron sequences are in *lower case letters*. The sequence of the 5′-flanking region is

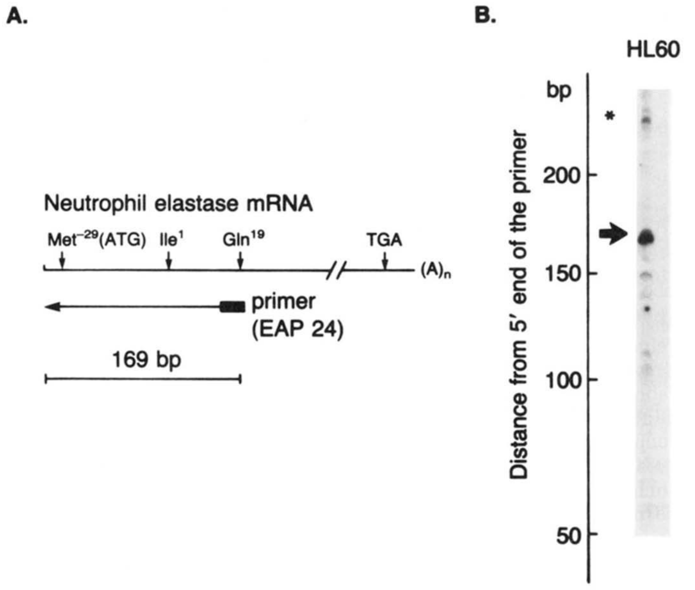

FIG. 4. **Identification of the major site of transcription initiation for the neutrophil elastase gene by primer extension analysis of HL60 mRNA.** *A*, structure of neutrophil elastase mRNA showing the position of the 5′ Met^{-29} (ATG), the Ile^{1} of the mature protein, the stop codon (TGA), and the poly(A) tail. The 5′ end of the primer (EAP 24) corresponds to the codon for Gln^{19}; the 3′ → 5′ polarity of the primer is indicated. *B*, primer extension evaluation of HL60 mRNA. A ^{32}P 5′ end-labeled synthetic oligonucleotide (5 ng) corresponding to the region of exon II underlined in Fig. 3 was annealed to poly(A)$^{+}$-selected RNA (10 μg) and extended by using reverse transcriptase. The reaction products were evaluated using an 8% polyacrylamide gel. The *arrow* indicates the likely major initiation point of transcription at a residue 26 bases upstream from the initiator ATG codon and 169 bases from the 5′ end of the primer. A faint band, indicated by the *asterisk*, is presented further upstream; this may identify a minor initiation site.

to that of the protein and partial cDNA sequence of a serine protease called "medullasin" isolated from human bone marrow (23) reveals that medullasin is actually human neutrophil elastase.

The sequences coding for the mature protein are spread out over exons II–V, with the sequence for the N-terminal Ile^{1} close to the 5′ end of exon II and the sequence for the C-terminal Gln^{218} in the middle of exon V. The three components of the catalytic active site His^{41}-Asp^{88}-Ser^{173} are found in three different exons (II, III, and V, respectively). Exon IV contains the sequences coding for the two *N*-linked glycosylation acceptor sites (Asn^{95}, Asn^{144}) that are used in the mature protein.

Beyond the Gln^{218} C terminus of the mature protein, the genomic sequence in exon V codes for a 20-residue "pro$_{C}$" extension. Unlike the mostly hydrophobic prepeptide, the pro$_{C}$ peptide is composed of a mixture of hydrophobic, acidic, and basic residues. Of interest, it contains a single cysteine residue and is proline-rich (5 of 20 residues).

Evaluation of subclone pPB74 encompassing the entire NE gene with a probe of Alu family sequences (BLUR 8) revealed sequences of this type in the region 5′ to exon I, in the intron between exons III and IV, and the region 3′ to exon V (Fig. 1). Evaluation of the sequence of the 5′-flanking region of exon I demonstrated a typical Alu family sequence between residues −577 and −218 (Fig. 3). In addition to the Alu-related sequence, the sequence in the region 5′ to exon I has a number of internal repeats (Figs. 3 and 5). In the region starting −1032 bp 5′ to exon I, there is a 317-bp sequence that contains six tandem repeats of 53 or 52 bp that are nearly identical. Interestingly, the 10-bp segment occupying the 3′ position of each of these repeats are also found at two locations (−468 and −338) within the Alu-related sequence (−577 to −218). Within the 1418 bp of the 5′-flanking region that was sequenced, there is also a pentamer, GGAGG, found 10 times, including twice in juxtaposition to the 10-bp repeat (−478, −348); four of these repeats occur within the Alu sequence (−577 to −218).

Comparison of the 5′-flanking region of the NE gene to other 5′-flanking regions revealed regions of homology to sequences 5′ to the myeloperoxidase (MPO) gene (24) and the c-*myc* gene (25) (Fig. 5*B*). A 19-base pyrimidine-rich (18 of 19 bases) sequence in the NE gene 5′-flanking region has 90% homology with an analogous region of the MPO gene. Although different from the region of homology between the NE and MPO genes, there are five regions of ≥75% homology between the NE and c-*myc* genes.

While the NE gene and its flanking regions contain a variety of repetitive sequences in the noncoding regions, the gene and its flanking regions are, as a whole, present in only one copy. In this regard, evaluation of gene copy number by dot hybridization using the NE clone λNE18.0 as a standard demonstrated that the NE gene was represented only once in the human haploid genome (Fig. 6). This result was supported by Southern blot analysis of human DNA using the partial NE cDNA pPB15 as a probe. When *Eco*RI-digested human DNA was hybridized with ^{32}P-labeled pPB15, a single 6.0-kb hybridization band was identified, identical in size to the *Eco*RI subclone pPB74 (data not shown). Consistent with these observations, *in situ* hybridization with ^{3}H-labeled NE cDNA pPB15 to chromosomal spreads prepared from normal

numbered from −1 to −1418 starting 5′ to the major transcription initiation "cap" site. For the 5′-flanking region, + indicates a 5-bp repeat and ♦ indicates a 53- or 52-bp repeat; ● indicates a 10-bp repeat within and outside the 53- or 52-bp repeat; and *hatched boxes* indicate the "CAAT," "TATA," and "GC" putative promoter sites. For the coding exons, amino acid numbers are shown on the *third line*. Negative numbers (−29 to −1) refer to the putative pre signal peptide (−29 to −3) and pro$_{N}$ peptide (−2 to −1). The N terminus of the mature protein (Ile^{1}) is in exon II, and the C terminus of the mature protein (Gln^{218}) is in exon V. There is a putative C-terminal "pro$_{C}$" precursor peptide of 20 residues (219–238). The region *underlined* in exon II signifies the complementary antisense oligonucleotide used in the primer extention study to identify the cap site (see Fig. 4). The single, in-frame start codon ATG is indicated as *start*, ⬇, the single in-frame stop codon TGA is indicated as *stop*, ⬆, and the putative polyadenylation signal (AATAAA) in the 3′-noncoding region is *underlined*. Exon-intron boundaries were determined by comparison of the genomic sequence with the NE cDNA sequence and NE RNA sequence. At the exon-intron junction boundaries where a coding exon is split, the corresponding amino acid is shown in *brackets*. *Open boxes* have been placed around the N-terminal Ile^{1} and the C-terminal Gln^{218} of the mature protein, *stippled boxes* around the two (Asn^{95}, Asn^{144}) *N*-linked oligosaccharide attachment sites used in the mature protein, and *hatched boxes* (His^{41}, Asp^{88}, Ser^{173}) represent the catalytic triad forming the active site. *Arrowhead* (▼) at residue Ser^{173} in exon V indicates a single base difference between the genomic DNA sequence (TC$\underline{\text{C}}$) and cDNA sequence (TC$\underline{\text{A}}$) (7) coding for the same amino acid.

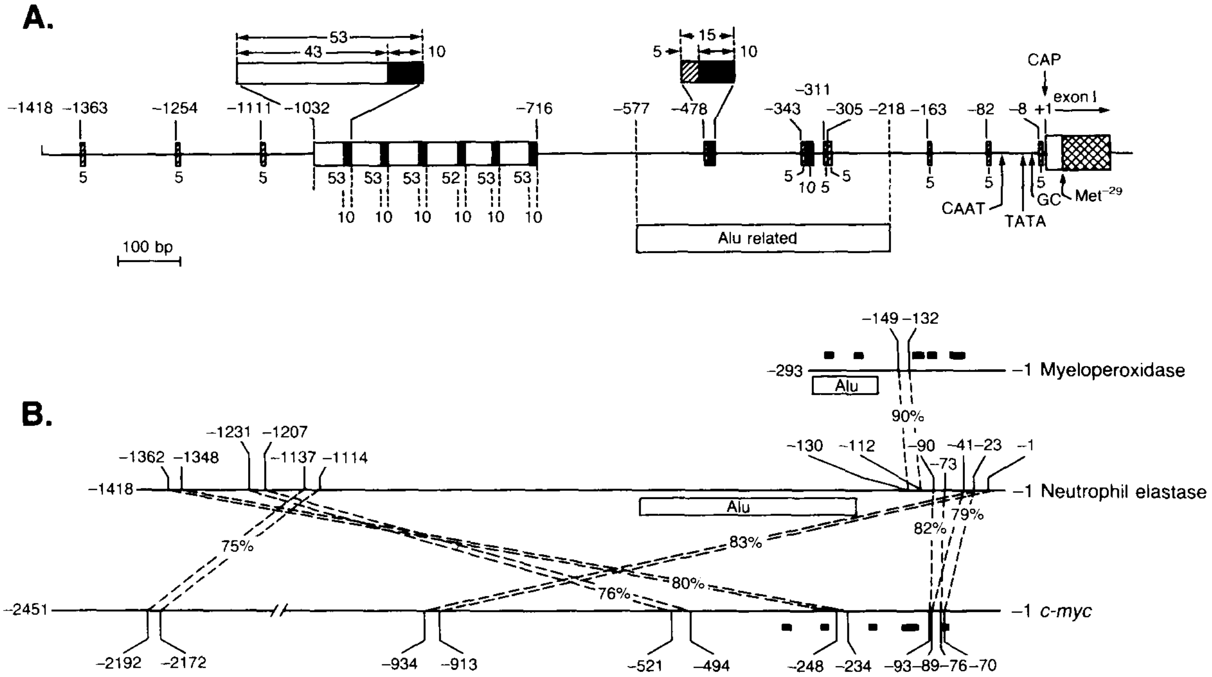

FIG. 5. **Structure of the 5′-flanking region of the neutrophil elastase gene.** *A*, overall structure of 1418 bp 5′ to the NE cap site (indicated by *+1, CAP*). The sequence of the region is shown in Fig. 3. Evaluation of the sequence demonstrates putative promoter consensus sites (CAAT, TATA, GC) within 61 bp 5′ to the cap site. Region −577 to −218 contains Alu-related sequences. Region −1032 to −716 contains six 53- or 52-bp repeats that include 10 bp (*black boxes*) which are observed elsewhere. There is also a 5-bp repeat (GGAGG, *hatched boxes*) present 10 times within the 1418-bp 5′-flanking region. *B*, homologies between the 5′-flanking regions of the NE gene, the myeloperoxidase (MPO) gene, and the c-*myc* gene, excluding Alu sequences (*open boxes*). The sequence for 293 bp of the MPO gene analyzed was that of Morishita *et al.* (24) and the 2451 bp of the c-*myc* gene that of Siebenlist *et al.* (25). Only regions of homology ≥75% are shown. The *numbers bracketing* each homologous region are indicated as minus base pairs from the cap site of the MPO gene and the second initiation site for transcription of the c-*myc* gene. The *black boxes above* the MPO sequence *and below* the c-*myc* sequence are the regions of homology between the 5′-flanking regions of MPO and c-*myc* observed by Morishita *et al.* (1987); note that the homologies between NE and MPO and those between NE and c-*myc* are different from the MPO-c-*myc* homologous regions.

human lymphocytes demonstrated a single location for the gene at q14 on chromosome 11 (Fig. 7).

DISCUSSION

Characterization of the gene for NE reveals it has many features common for the serine group of proteases (26), but with some components apparently specific for NE. Like most serine proteases, the NE gene sequence predicts a precursor protein with the signal and pro$_N$ peptides. In addition, the NE gene predicts a pro$_C$ peptide. This predicted structure has potential relevance to the evolution, targeting, and activation of this protease.

Evolution—Mammalian serine proteases are thought to have derived from a common ancestor by divergent evolution (27). In this context, evaluation of the primary sequence and three-dimensional crystallographic structure of NE demonstrates a number of features typical of the serine protease family, particularly the conservation of the His-Asp-Ser catalytic site (4). Consistent with this concept, the structure of the NE gene shows a number of features typical of this family, including: 1) each of the three components of the catalytic triad is coded by sequences in a different exon, and, for each, the codon for the active site amino acids is within 18 bp from the 3′ or 5′ end of the exon (28, 29); 2) the coding exons predict a precursor protein with the signal and pro$_N$ peptides (29, 30); and 3) the sequence for the protease domain is spread over several exons (4 for NE), and the introns between the exons coding for the protease domain follow a consistent pattern of location and phase (28, 29). Interestingly, comparison of the gene structure of NE to that of the other serine proteases localized to cells derived from bone marrow in which the gene sequence is known (cathepsin G, mast cell protease), shows an even more remarkable similarity than among the serine proteases in general (31, 32). For this subgroup of serine proteases, the pro$_N$ peptide is always 2 amino acids with the sequence *X*-Glu. In contrast, the pro$_N$ peptide of the pancreatic cell serine proteases are 6–20 residues in length and end in Arg or Lys (28, 33), and many other serine proteases have even longer pro$_N$ peptides and many end in Arg (30). In addition, the introns of mast cell protease are identical in location and phase to those of NE (32). Consistent with the concept that the serine proteases derived by divergent evolution, they are widely dispersed throughout the genome. In this regard, the NE gene is localized at 11q14. The prothrombin gene is the closest known serine protease, mapping at 11p11-q12 (34).

Targeting—In the granulocytic series, NE is found in the neutrophil and its progenitors in bone marrow (35). After synthesis, NE is eventually shunted into so-called "azurophilic" or "primary" granules, 0.5 μm of lysosomal-like structures found in the cytoplasm of granulocyte precursors from the promyelocyte stage onward (35, 36). The mechanism by which NE is targeted to these granules is unknown. In the context that azurophilic granules are lysosomal-like, it is reasonable to assume that the targeting may be similar to lysosomal enzymes in general. In this regard, if one of the

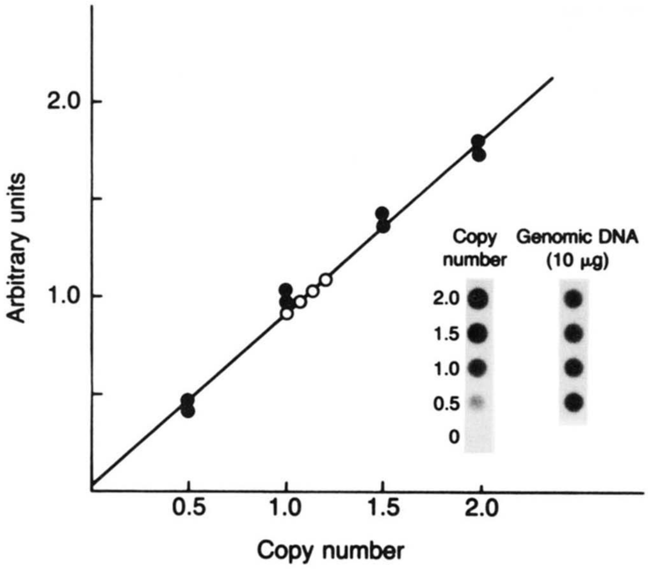

FIG. 6. **Determination of the copy number of neutrophil elastase gene in human genomic DNA using a ^{32}P-labeled cDNA (pPB15) as a probe.** Total human genomic DNA digested with *Eco*RI was evaluated using the NE genomic clone λNE18.0 as a standard (equivalent to 0, 0.5, 1.0, 1.5, 2.0 gene copies). The autoradiographic images of the dot hybridizations were scanned, and the value for zero copies subtracted from the other values. The absorbance of the dots is displayed (in arbitrary units, on the *ordinate*) *versus* the number of gene copy equivalents present in the dots (on the *abscissa*). *Closed circles* represent absorbances of the standard curve of gene copy equivalent dots of the NE clone λNE 18.0. *Open circles* represent absorbance of the total human genomic DNA (10 μg/dot). *Inset*, example of the dot blot analysis for standard curve (*left*) and genomic DNA (*right*). The standard curve was fitted using the method of least squares.

carbohydrate side chains of the mature NE contains terminal mannose residues, it is possible that, like other lysosomal enzymes, mannose residues are phosphorylated in the Golgi, recognized by a mannose 6-phosphate receptor, and the NE shunted to lysosomes (37).

In the context that the NE gene predicts that the protein is produced in a precursor form, it is conceivable that the pre, pro$_N$, or pro$_C$ peptides play some role in the targeting of the enzyme. The 27 N-terminal residues predicted by the NE gene sequence are so typical of a pre signal peptide, it is likely that this peptide targets the newly synthesized protein to the endoplasmic reticulum and that this peptide is removed in the process (20–22). Possible roles for the pro$_N$ and/or pro$_C$ pieces in the targeting process, if any, are only speculative. Comparison of the predicted NE precursor sequence to other components of the neutrophil azurophilic granules in which the precursor sequence is known in whole or in part (*e.g.* myeloperoxidase, cathepsin G, cathepsin D, and β-glucuronidase) does not give any direct clues to how the targeting occurs (24, 31, 38, 39). It is of interest, however, that the N-terminal sequences of preproinsulin can direct the translation and glycosylation of an irrelevant protein by mammalian microsomal membranes (40), *i.e.* the N- and/or C-terminal precursor peptides may be responsible for targeting of NE.

Activation—Like other proteases, NE is a potential hazard to the cell producing it, and thus it is reasonable to assume that the primary translation product, even without the "pre" sequence, is inactive. Although the general structure for NE predicts a primary translation product of 267 residues of the form pre-pro$_N$-NE-pro$_C$, the amino acid sequence of the protein found in azurophilic granules is the mature form of NE, it is likely that the NE is activated prior to, or during, storage in the azurophilic granules. For most serine proteases, the "activation" occurs after cleavage of an N-terminal pro sequence of the inactive zymogen (27). Such N-terminal "activation" peptides can be of varying length, as short as 6 amino acids for bovine trypsinogen (6) and as much as 560 amino acids for plasminogen (27), and thus it is possible that the pro$_N$ and/or the 20-residue pro$_C$ peptide have such a function. In this regard, cathepsin G, another serine protease stored in the neutrophil azurophilic granules, has a predicted pro$_C$ peptide.[2] In addition or alternatively to a role in activation, the pro$_N$ and/or pro$_C$ peptides of the NE precursor may have other functions. In this regard, the pro$_N$ peptide of carboxypeptidase A inactivates carboxypeptidase A itself (41).

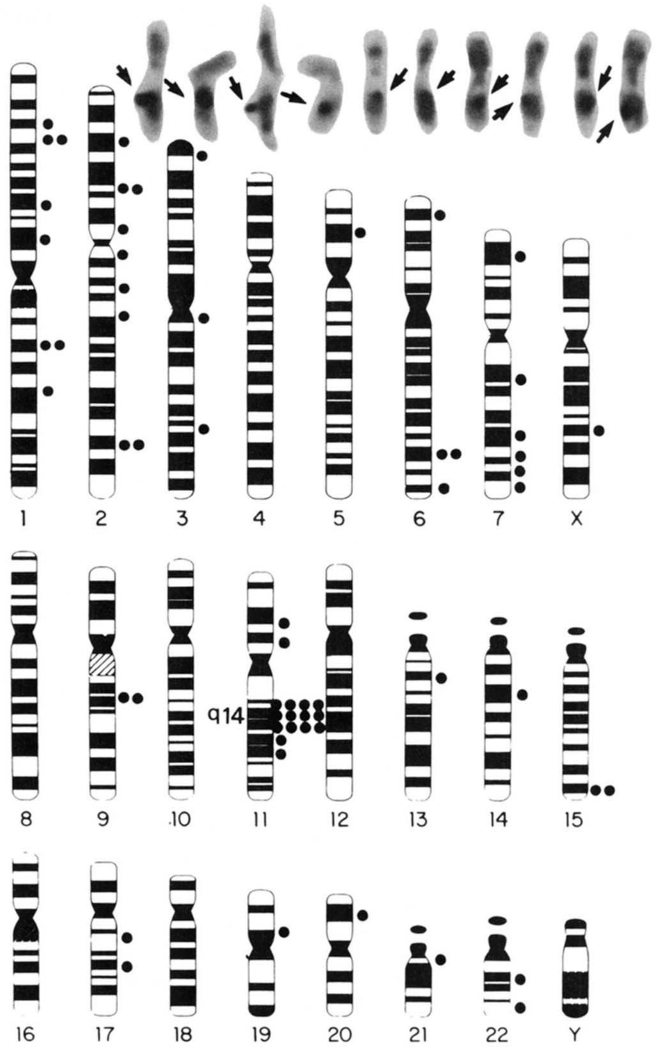

FIG. 7. **Chromosomal localization of the human neutrophil elastase gene.** Shown is a schematic representation of fluorodeoxyuridine-synchronized Wright-stained metaphase chromosomes depicting all chromosomal grains observed in 40 cells evaluated by *in situ* hybridization with a ^{3}H-labeled, 0.65-kb insert of cDNA clone (pPB15). Of the total grains observed, 20% were localized to band 11q14. In the *upper part* of the figure is shown 10 examples of chromosome 11 from different metaphases, with *arrows* indicating the individual grains.

Modulation of NE Gene Expression—Despite its name, NE is not synthesized by neutrophils, as evidenced by the observation that mature blood neutrophils contain no detectable NE mRNA transcripts (7). However, human bone marrow precursors contain NE mRNA transcripts and studies with

[2] J. Travis, personal communication.

the HL60 tumor cell line have shown that when stimulated to differentiate toward the granulocytic series, NE mRNA levels increase, and when stimulated toward the monocytic series, NE mRNA transcripts decrease. Together, these observations suggest the NE gene is activated in promyelocytes, myelocytes, and metamyelocytes but is shut down prior to the departure of the mature neutrophil from the marrow.

Evaluation of the DNA sequence 5′ to exon I of the NE gene reveals several interesting segments that may be relevant to NE gene expression. First, there is remarkable homology of the 19-base pyrimidine-rich (18 of 19 bases) sequence with the 5′-flanking region of the MPO gene, an enzyme found in the same cell types and also stored in the azurophilic granules (24). However, the MPO gene is likely activated earlier and shut off earlier in myelopoiesis than in NE (42). Furthermore, while activation of HL60 cells with dimethyl sulfoxide upregulates the NE gene, it suppresses the MPO gene (42). Thus, if this region of homology bears relevance to the activation of these genes, the mechanisms involved are likely complex.

Second, the region 5′ to the NE gene also contains several regions of homology to the region 5′ to the proto-oncogene c-*myc* (25). Homology between the regions 5′ to MPO and c-*myc* have also been noted (24), although such segments are all different from the NE-c-*myc* homologous regions. Interestingly, when HL60 cells are stimulated with either dimethyl sulfoxide or PMA, c-*myc* mRNA transcripts rapidly disappear (43), while dimethyl sulfoxide increases the numbers of NE transcripts and PMA has the opposite effect. Thus, like the NE-MPO 5′ homologies, the relevance of the NE-c-*myc* homologies, if any, is presently obscure.

Third, in addition to these homologies, the 5′ region of the NE gene contains a variety of repetitive sequences. Among these, the 6-fold tandem repeats of 53 or 52 bp represented are of interest, although their function is unknown. For example, the 21- and 72-bp tandem repeats of the SV40 promoter function as enhancer elements binding gene activation proteins (44, 45). In addition, functional repetitive elements have been identified in other eukaryotic genes including those of humans (46–48). The 5′ region also contains 10 repeats of the pentamer GGAGG. While any pentamer sequence would be expected once every 1024 bp on a theoretical basis, the GGAGG repeat occurs 10 times in the 1418-bp sequence immediately 5′ to exon I of the NE gene, *i.e.* an average of once every 142 bp, a frequency more than 7 times expected.

Finally, while all of the available evidence suggests the cap site 26 bp 5′ to the ATG is the major site of NE gene transcription, there may be a minor transcription initiation site approximately 90 bp 5′ to the ATG. Assuming a typical 200–250-base poly(A) tail, both initiation sites are compatible with 1.3-kb size NE mRNA identified by Northern blot analysis (7). Interestingly, while there are no typical TATA or GC boxes in association with this second site, there is a CAAT box-like sequence in the region starting at −150 and a TATA-like box sequence in the region starting at −624. Thus, it is conceivable there may be an additional 5′ exon(s) and associated controlling elements for this gene.

Acknowledgment—We would like to thank L. Sichert for excellent editorial assistance.

REFERENCES

1. Janoff, A., and Scherer, J. (1968) *J. Exp. Med.* **128,** 1137–1155
2. Sinha, S., Watorek, W., Karr, S., Giles, J., Bode, W., and Travis, J. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 2228–2232
3. Bieth, J. G. (1986) in *Regulation of Matrix Accumulation* (Mecham, R., ed) pp. 217–320, Academic Press, New York
4. Bode, W., Wei, A. Z., Huber, R., Meyer, E., Travis, J., and Neumann, S. (1986) *EMBO J.* **5,** 2453–2458
5. Janoff, A. (1985) *Ann. Rev. Med.* **36,** 207–216
6. Huber, R., and Bode, W. (1977) *Acc. Chem. Res.* **11,** 114–122
7. Takahashi, H., Nukiwa, T., Basset, P., and Crystal, R. G. (1988) *J. Biol. Chem.* **263,** 2543–2547
8. Ganz, T., Metcalf, J. A., Gallin, J. I., and Hehrer, R. I. (1987) *Clin. Res.* **35,** 424A
9. Gadek, J. E., and Crystal, R. G. (1982) in *Metabolic Basis of Inherited Disease* (Stanbury, J. B., Wyngaarden, J. B., Fredrickson, D. S., and Brown, M. S., eds) pp. 1450–1467, McGraw-Hill, New York
10. Nukiwa, T., Satoh, K., Brantly, M. L., Ogushi, F., Fells, G. A., Courtney, M., and Crystal, R. G. (1986) *J. Biol. Chem.* **261,** 15989–15994
11. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
12. Tabor, S., and Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 4767–4771
13. Geliebter, J., Zeff, R. A., Melvold, R. W., and Nathenson, S. G. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 3371–3375
14. Giaever, G., Lynn, R., Goto, T., and Wang, J. C. (1986) *J. Biol. Chem.* **261,** 12448–12454
15. Rigby, P. W. J., Dieckmann, M., Rhodes, C., and Berg, P. (1977) *J. Mol. Biol.* **113,** 237–251
16. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual,* pp. 97–148, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
17. Deininger, P. L., Jolly, D. J., Rubin, C. M., Friedmann, T., and Schmid, C. W. (1981) *J. Mol. Biol.* **151,** 17–33
18. Dalgleish, R., Trapnell, B. C., Crystal, R. G., and Tolstoshev, P. (1982) *J. Biol. Chem.* **257,** 13816–13822
19. Merlino, G. T., Ishii, S., Whang-Peng, J., Knutsen, T., Xu, Y. H., Clark, A. J., Stratton, R. H., Wilson, R. K., Ma, D. P., Roe, B. A., Hunts, J. H., Shimizu, H., and Pastan, I. (1985) *Mol. Cell Biol.* **5,** 1722–1734
20. von Heijne, G. (1983) *Eur. J. Biochem.* **133,** 17–21
21. Watson, M. E. E. (1984) *Nucleic Acids Res* **12,** 5145–5164
22. Perlman, D., and Halvorson, H. O. (1983) *J. Mol. Biol.* **167,** 391–409
23. Okano, K., Aoki, Y., Sakurai, T., Kajitani, M., Kanai, S., Shimazu, T., Shimizu, H., and Naruto, M. (1987) *J. Biochem. (Tokyo)* **102,** 13–16
24. Morishita, K., Tsuchiya, M., Asano, S., Kaziro, Y., and Nagata, S. (1987) *J. Biol. Chem.* **262,** 15208–15213
25. Siebenlist, U., Hennighausen, L., Battey, J., and Leder, P. (1984) *Cell* **37,** 381–391
26. Kraut, J. (1977) *Annu. Rev. Biochem.* **46,** 331–358
27. Neurath, H. (1984) *Science* **224,** 350–357
28. Craik, C. S., Rutter, W. J., and Fletterick, R. (1983) *Science* **220,** 1125–1129
29. Rogers, J. (1985) *Nature* **315,** 458–459
30. Asakai, R., Davie, E. W., and Chung, D. W. (1987) *Biochemistry* **26,** 7221–7228
31. Salvesen, G., Farley, D., Shuman, J., Przybyla, A., Reilly, C., and Travis, J. (1987) *Biochemistry* **26,** 2289–2293
32. Benfey, P. N., Yin, F. H., and Leder, P. (1987) *J. Biol. Chem.* **262,** 5377–5384
33. Isackson, P. J., Ullrich, A., and Bradshaw, R. A. (1984) *Biochemistry* **23,** 5997–6002
34. Royle, N. J., Irwin, D. M., Koschinsky, M. L., MacGillivray, R. T., and Hamerton, J. L. (1987) *Somatic Cell Mol. Genet.* **13,** 285–292
35. Bainton, D. F., Ullyot, J. L., and Farquhar, M. G. (1971) *J. Exp. Med.* **134,** 907–934
36. Falloon, J., and Gallin, J. I. (1986) *J. Allergy. Clin. Immunol.* **77,** 653–662
37. Kornfeld, S. (1986) *J. Clin. Invest.* **77,** 1–6
38. Faust, P. L., Kornfeld, S., and Chirgwin, J. M. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 4910–4914
39. Oshima, A., Kyle, J. W., Miller, R. D., Hoffmann, J. W., Powell, P. P., Grubb, J. H., Sly, W. S., Tropak, M., Guise, K. S., and Gravel, R. A. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 685–689

40. Eskridge, E. M., and Shields, D. (1986) *J. Cell Biol.* **103,** 2263-2272

41. Segundo, B. S., Martinez, M. C., Vilanova, M., Cuchillo, C. M., and Aviles, F. X. (1982) *Biochim. Biophys. Acta.* **707,** 74-80

42. Weil, S. C., Rosner, G. L., Reid, M. S., Chisholm, R. L., Farber, N. M., Spitznagel, J. K., and Swanson, M. S. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 2057-2061

43. Davis, R. C., Thomason, A. R., Fuller, M. L., Slovin, J. P., Chou, C. C., Chada, S., Gatti, R. A., and Salser, W. A. (1987) *Dev. Biol.* **119,** 164-174

44. Gruss, P., Dhar, R., and Khoury, G. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 943-947

45. Banerji, J., Rusconi, S., and Schaffner, W. (1981) *Cell* **27** 299-308

46. Südhof, T. C., Van Der Westhuyzen, D. R., Goldstein, J. L., Brown, M. S., and Russell, D. W. (1987) *J. Biol. Chem.* **262,** 10773-10779

47. Kuhl, D., de la Fuente, J., Chaturvedi, M., Parimoo, S., Ryals, J., Meyer, F., and Weissmann, C. (1987) *Cell* **50,** 1057-1069

48. Deutsch, P. J., Jameson, J. L., and Habener, J. F. (1987) *J. Biol. Chem.* **262,** 12169-12174